TWO TYPES OF HAPAXOMERS

- INTERNAL PALINDROMIC TYPE II ENZYMES
  -e.g., Sfi I

  G G C C N N N N^N G G C C

  C C G G N^N N N N C C G G

- OUTSIDE CUTTERS (TYPE IIS)
  -e.g., Sap I

  G C T C T T C N^N N N

  C G A G A A G N N N N^

FIG. 1

CAGNNNCTG

CAGNNNCTG

FIG. 2A

GGATGNNNNNNNNNNNNNN

CCTACNNNNNNNNNNNNNN

NNNNNNNNNNNNNNCATCC

NNNNNNNNNNNNNNGTAGG

FIG. 2B

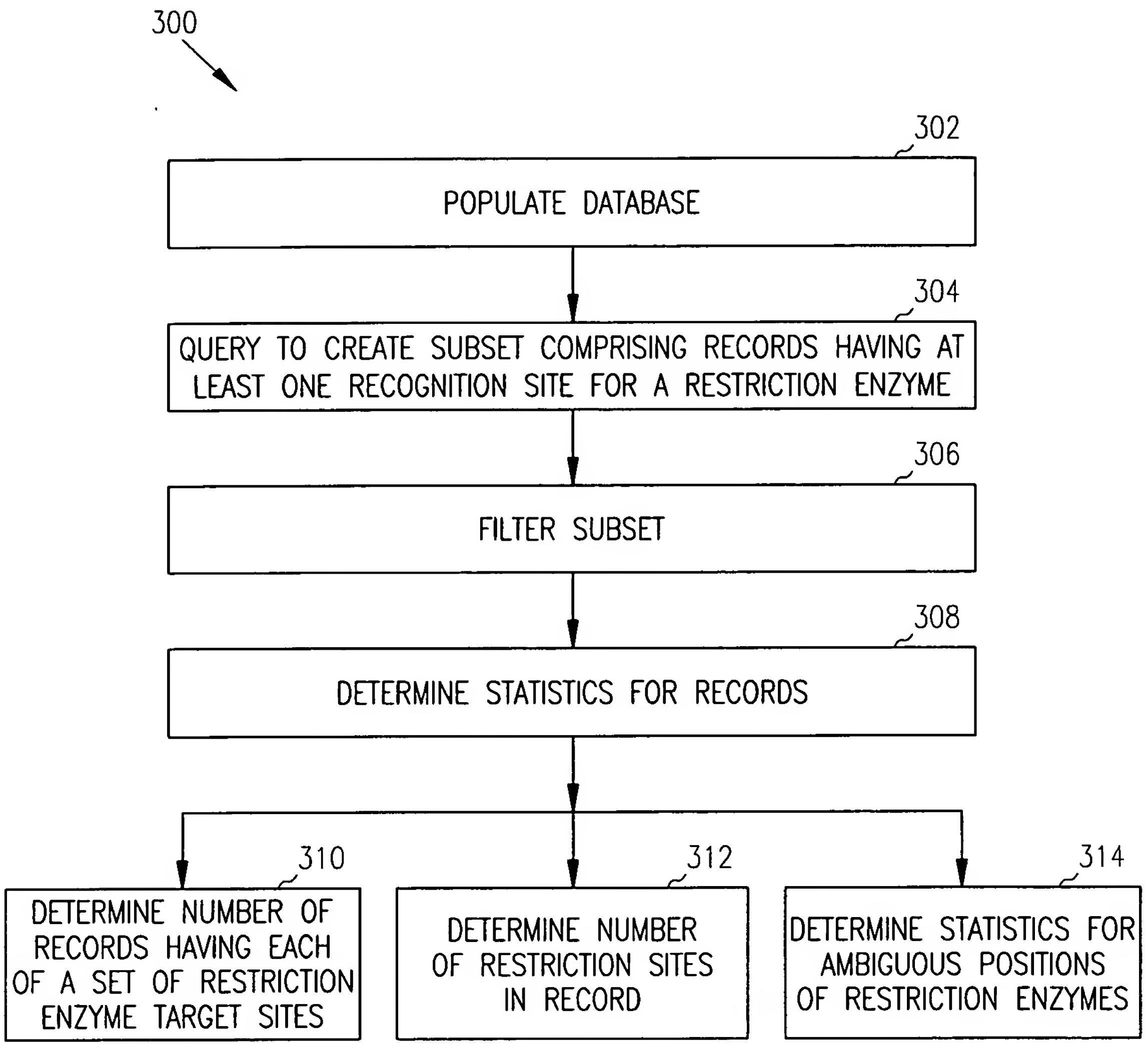
300
302
POPULATE DATABASE
304
QUERY TO CREATE SUBSET COMPRISING RECORDS HAVING AT LEAST ONE RECOGNITION SITE FOR A RESTRICTION ENZYME
306
FILTER SUBSET
308
DETERMINE STATISTICS FOR RECORDS
310
DETERMINE NUMBER OF RECORDS HAVING EACH OF A SET OF RESTRICTION ENZYME TARGET SITES
312
DETERMINE NUMBER OF RESTRICTION SITES IN RECORD
314
DETERMINE STATISTICS FOR AMBIGUOUS POSITIONS OF RESTRICTION ENZYMES

FIG. 3

	Sap I (0)	Sfi I (0)	Sfi I (0+1)	Sfi (0-2)	Sgf I/Pme I	Sap I (0)	Sap I (0+1)	Sap I (0-2)
Hs_Fna	56.72	85.93	96.72	98.91	97.45	56.72	84.67	94.82
Mgc	61.98	87.02	97.94	99.70	98.14	61.98	90.31	97.60
E coli	86.71	99.35	100.00	100.00	93.52	86.71	98.37	99.53
C elegans	65.93	99.46	99.97	100.00	98.55	65.93	90.38	97.19
S cerevisiae	80.02	99.15	99.98	100.00	97.17	80.02	96.68	99.29
Arabidopsis	70.83	99.56	100.00	100.00	97.57	70.83	93.37	98.63

FIG. 4

7+ Cutters

Enzymes	Recognition Sequence	HsFna	MGC	Ec	Ce	Sc	At
AarI	CACCTGCNNNN^NNNN	7142	5355				1138
AbeI	*CC^TCA_GC not available*	*7970*	*5836*	*141*	*90*	*374*	*1833*
AscI	GG^CGCG_CC	515	336	152	10	13	26
AsiSI	GCG_AT^CGC	108	62	207	39	29	178
BbvCI	*CC^TCA_GC*	*7970*	*5836*	*141*	*90*	*374*	*1833*
CciNI	GC^GGCC_GC	1444	823	19	33	31	97
CpoI	*CG^GWC_CG*	*1119*	*781*	*347*			
CspI	*CG^GWC_CG*	*1119*	*781*	*347*			
CspBI	GC^GGCC_GC not available	1444	823	19	33	31	97
FseI	GG_CCGG^CC	1139	740	5	9	10	70
MabI	*A^CCWGG_T*						
MchAI	GC^GGCC_GC not available	1444	823	19	33	31	97
Mlu1106I	*RGGWCCY* not available						
NotI	GC^GGCC_GC	1444	823	19	33	31	97
PacI	TTA_AT^TAA	708	395	66	8	213	138
Pfl27I	*RG^GWC_CY* not available						
PpuMI	*RG^GWC_CY*						
PpuXI	*RG^GWC_CY*						
Psp5II	*RG^GWC_CY*						
PspPPI	*RG^GWC_CY*						
RsrII	*CG^GWC_CG*	*1119*	*781*	*347*			
Rsr2I	*CG^GWC_CG*	*1119*	*781*	*347*			
SanDI	*GG^GWC_CC*						
SapI	GCTCTTCN^NNN_	7260	4785	584	1296	1362	8870
SbfI	CC_TGCA^GG	2591	1802	60	13	66	251
SdaI	CC_TGCA^GG	2591	1802	60	13	66	251
SdiI	GGCCN_NNN^NGGCC not available	2214	1634	28	18	54	121
SexAI	*A^CCWGG_T*						
SfiI	GGCCN_NNN^NGGCC	2214	1634	28	18	54	121
SgfI	GCG_AT^CGC	108	62	207	39	29	178
SgrAI	CR^CCGG_YG						
Sse232I	CG^CCGG_CG not available	708	448	29	43	23	446
Sse1825I	*GG^GWC_CC* not available						
Sse8387I	CC_TGCA^GG	2591	1802	60	13	66	251
Sse8647I	*AG^GWC_CT* not available						
VpaK32I	GCTCTTCN^NNN_ not available	7260	4785	584	1296	1362	8870

FIG. 5A

Six Cutters

Enzymes	Recognition Sequence	HsFna	MGC	Ec	Ce	Sc	At
NruI	TCG^CGA	830	607	1070	558	507	2422
SplI	C^GTAC_G	701	449	498	263	549	1705
SnaBI	TAC^GTA	1080	621	435	165	885	2164
PvuI	CG_AT^CG	842	512	1000	705	537	3078
MluI	A^CGCG_T	1049	1019	976	295	337	1824
BglI	GCCN_NNN^NGGC	8827	6868	1333	469	750	2576
Ear I	CTCTTCN^NNN_						
BsrGI	TGTACA	6683	4551	442	760	1583	5450
XmnI	GAANN^NNTTC	8401	5850	1167	1652	2911	12141
SalI	G^TCGA_C	1515	944	463	792	856	3616
BamHI	G^GATC_C	6426	4305	438	1047	1238	6782
KpnI	G_GTAC^C	4098	2755	442	305	1317	2992
EcoRI	G^AATT_C	6536	4132	470	1346	2466	8244
XhoI	C^TCGA_G	3651	2402	156	800	737	7092
EcoRV	GAT^ATC	3789	2435	1378	919	2289	8419
	AAA^TTT	7484	5167	1008	1049	3843	8078
DraI	TTT^AAA	8455	6243	967	494	3018	6778

FIG. 5B

7+ Blunt Cutters

Enzymes	Recognition Sequence	HsFna	MGC	Ec	Ce	Sc	At
BstRZ246I	ATTT^AAAT	1204	648	55	38	379	317
BstSWI	ATTT^AAAT	1204	648	55	38	379	317
MspSWI	ATTT^AAAT	1204	648	55	38	379	317
MssI	GTTT^AAAC	297	173	71	9	152	490
PmeI	GTTT^AAAC	297	173	71	9	152	490
SmiI	ATTT^AAAT	1204	648	55	38	379	317
SwaI	ATTT^AAAT	1204	648	55	38	379	317
SrfI	GCCC^GGGC	1433	887	40	11	11	30

FIG. 5C

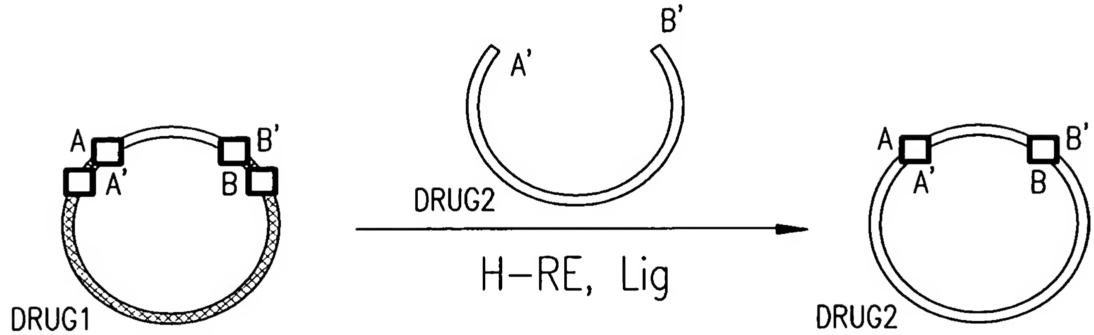

FIG. 6A

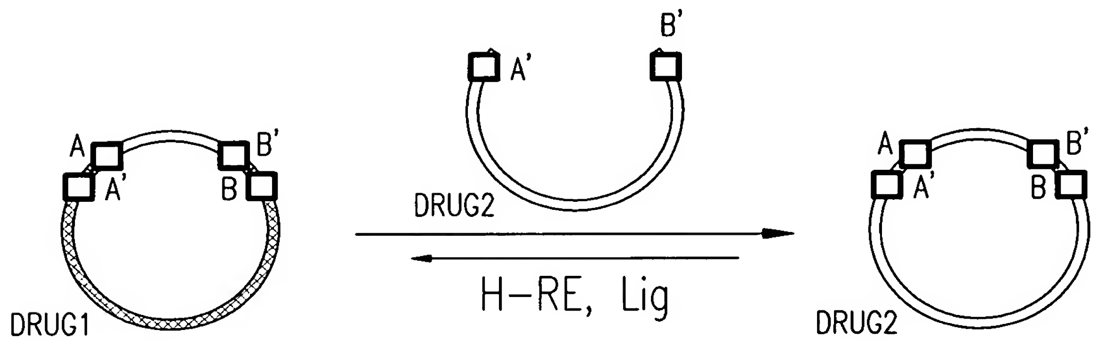

FIG. 6B

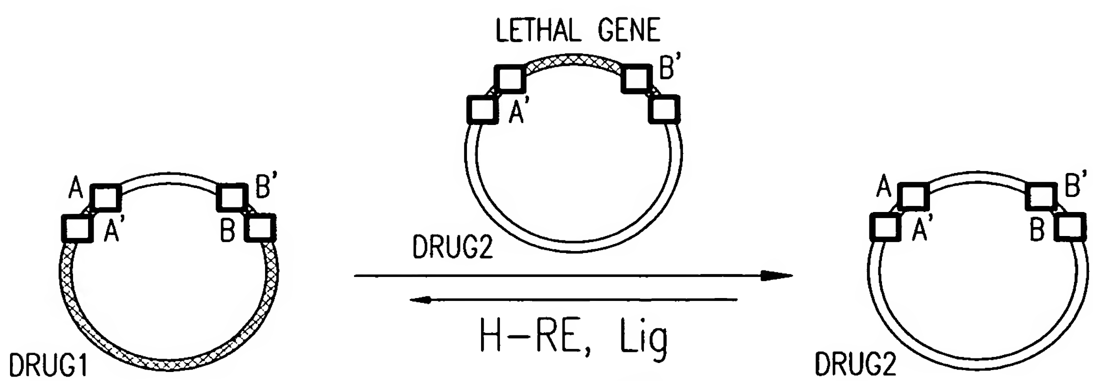

FIG. 6B

Sfi

- HOW TO MAKE Sfi I "ONE WAY"
  - -METHYLASES
  - -Bgl, NOT Sfi I SITS, IN ACCEPTOR VECTORS

```
G G C C N N N N^N G G C C
C C G G N^N N N N C C G G

  G C C N N N N^N G G C
  C G G N^N N N N C C G
```

  - -LETHAL GENES IN STUFFER FRAGMENTS

FIG. 7

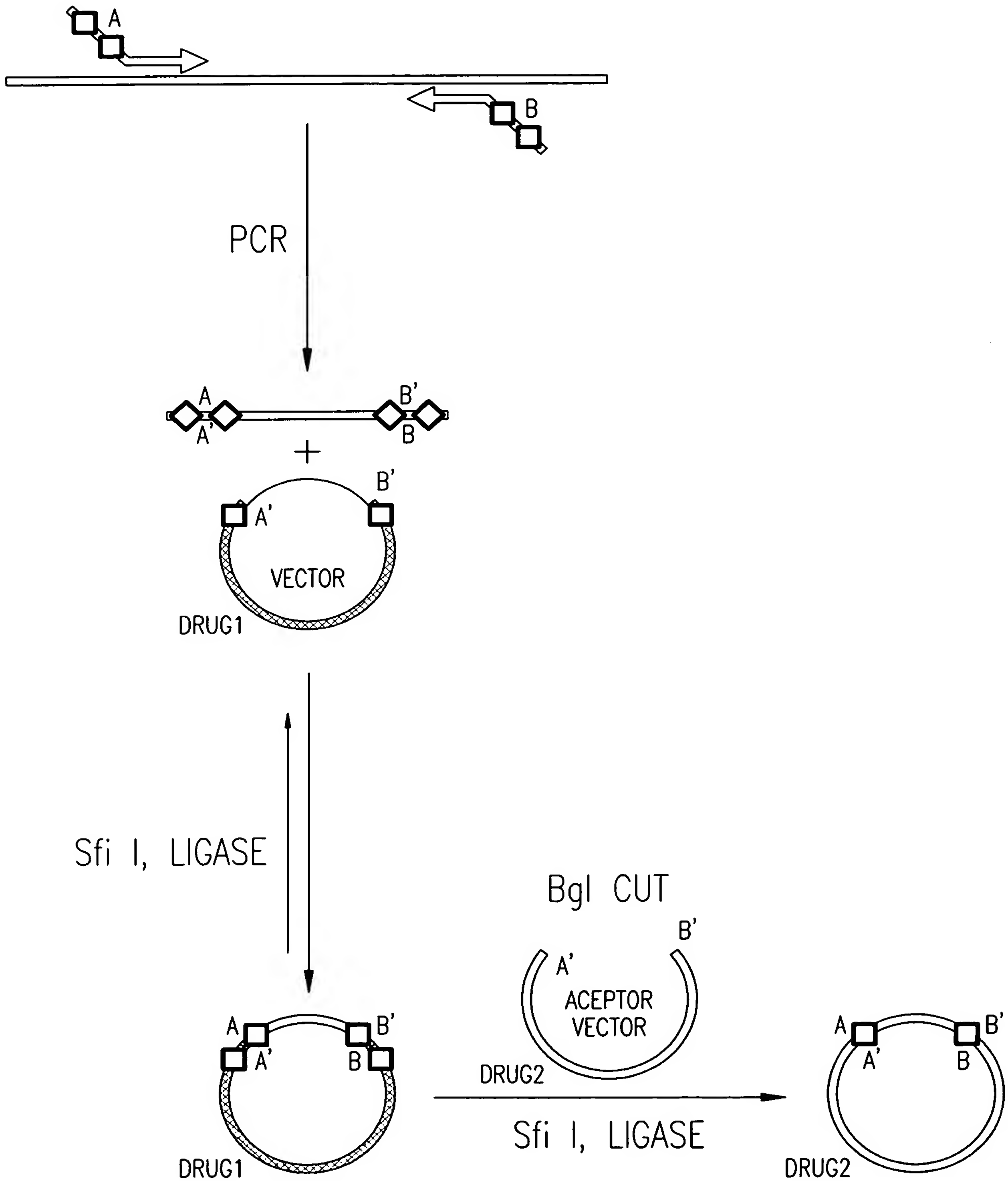
A
B
PCR
A
A'
B'
B
+
B'
A'
VECTOR
DRUG1
Sfi I, LIGASE
Bgl CUT
B'
A'
ACEPTOR
VECTOR
DRUG2
A
A'
B'
B
DRUG1
Sfi I, LIGASE
A
A'
B'
B
DRUG2

FIG. 8

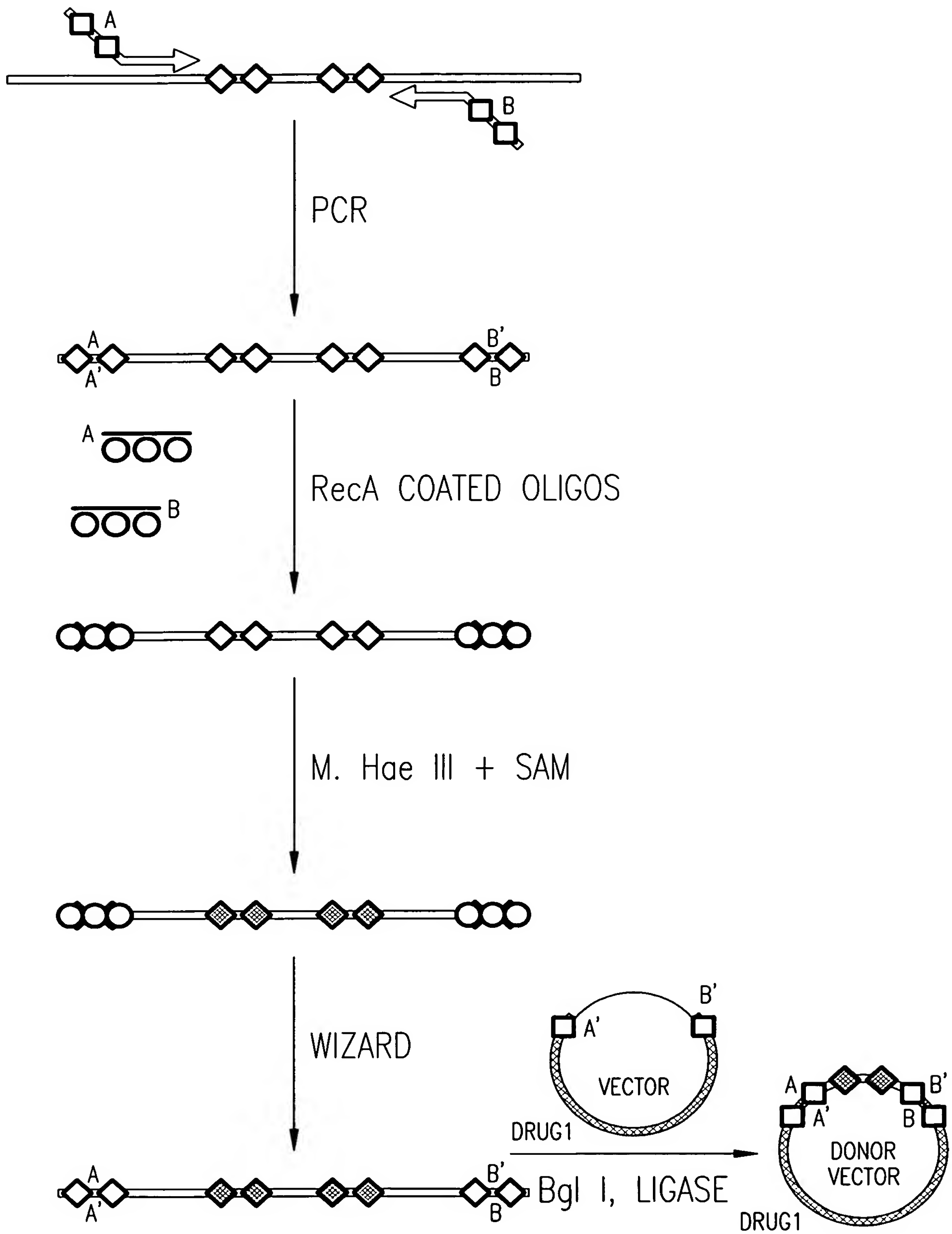
A
B
PCR
A
A'
B'
B
A
B
RecA COATED OLIGOS
M. Hae III + SAM
WIZARD
B'
A'
VECTOR
DRUG1
A
A'
B'
B
DRUG1
Bgl I, LIGASE
A
A'
B'
B
DONOR
VECTOR
DRUG1

FIG. 9A

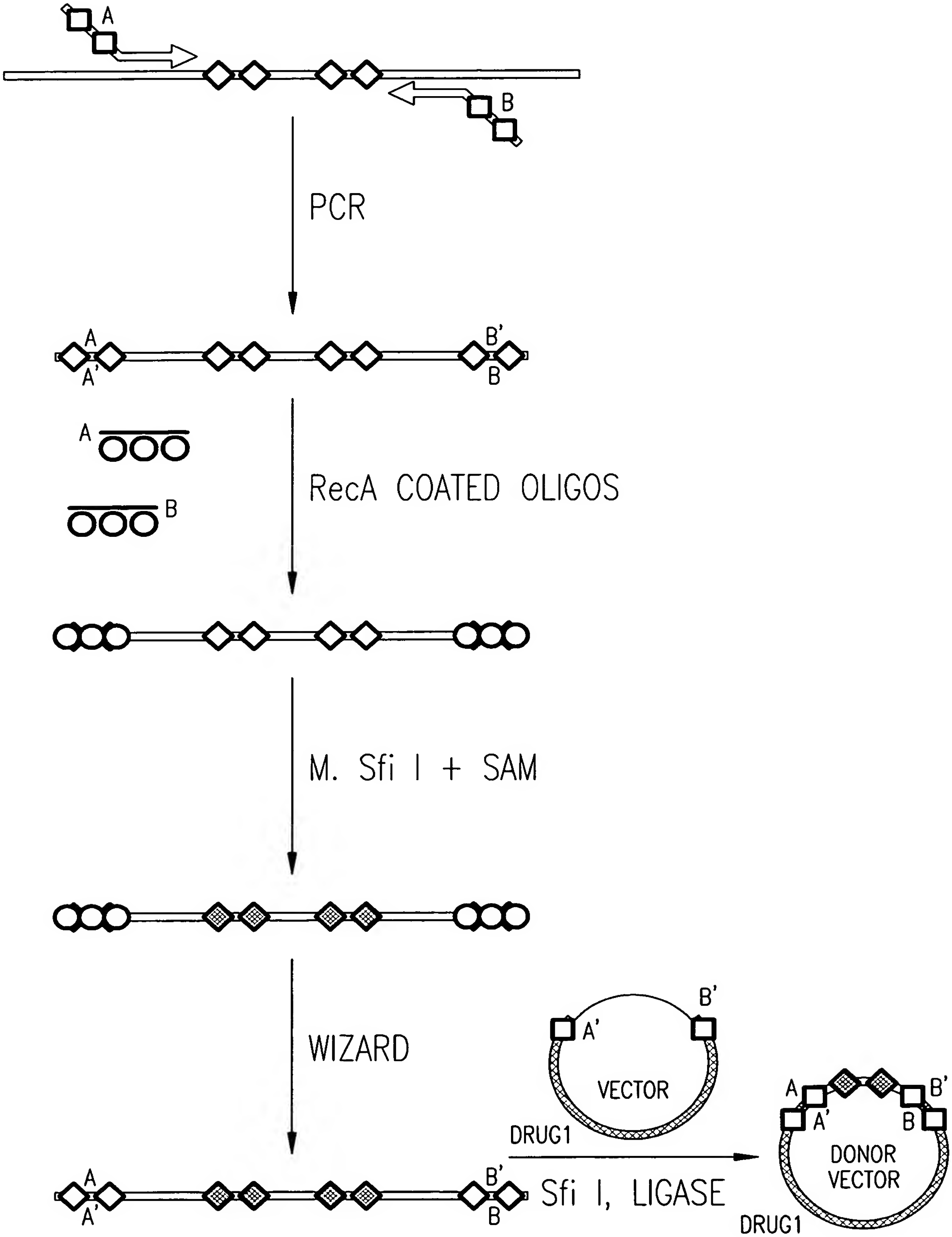
A
B
PCR
A
A'
B'
B
A
B
RecA COATED OLIGOS
M. Sfi I + SAM
WIZARD
A
A'
B'
B
B'
A'
VECTOR
DRUG1
Sfi I, LIGASE
A
B'
A'
B
DONOR
VECTOR
DRUG1

FIG. 9B

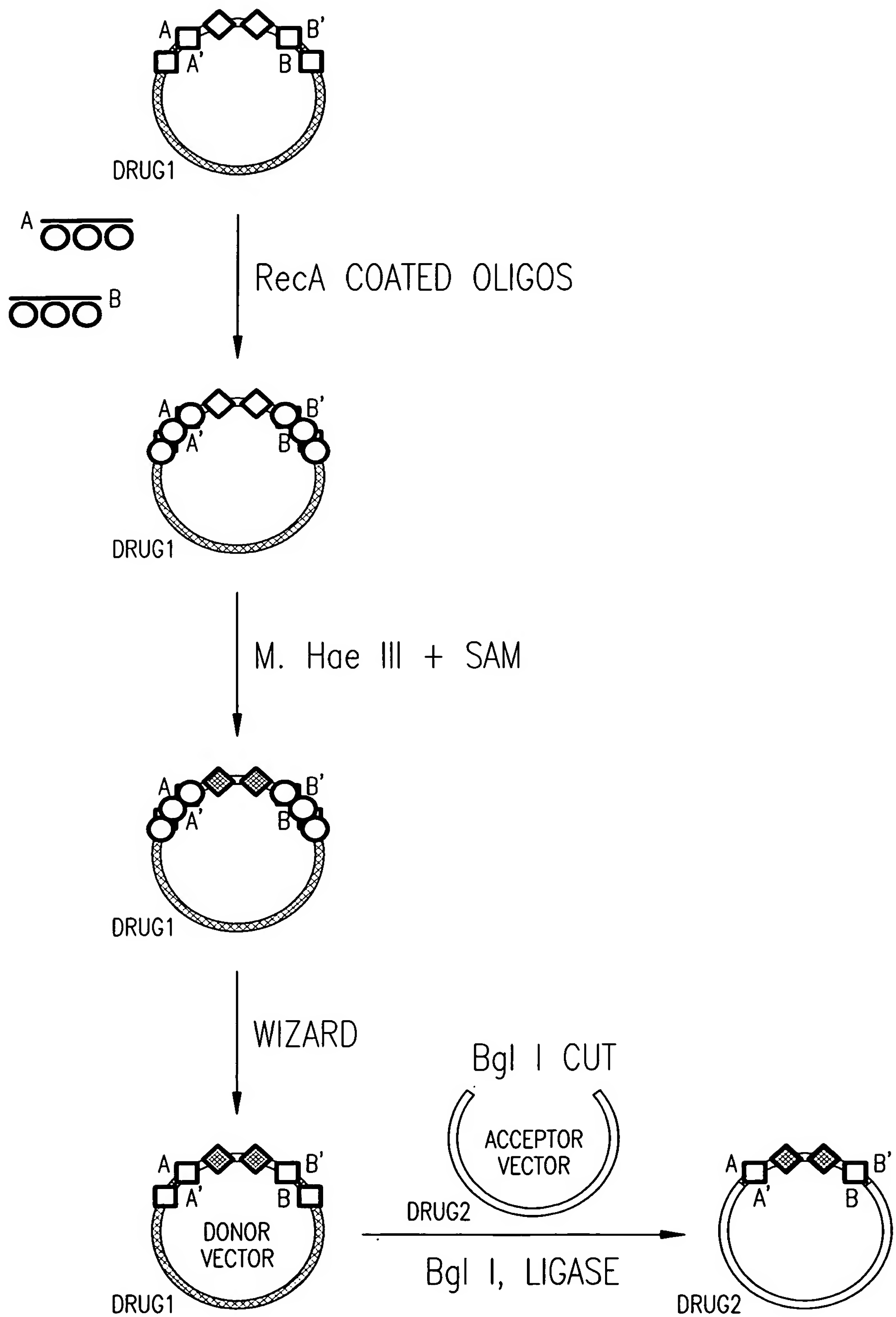
A
A'
B
B'
DRUG1
A
B
RecA COATED OLIGOS
A
A'
B
B'
DRUG1
M. Hae III + SAM
A
A'
B
B'
DRUG1
WIZARD
Bgl I CUT
ACCEPTOR
VECTOR
DRUG2
A
A'
B
B'
DONOR
VECTOR
DRUG1
Bgl I, LIGASE
A
A'
B
B'
DRUG2

FIG. 10A

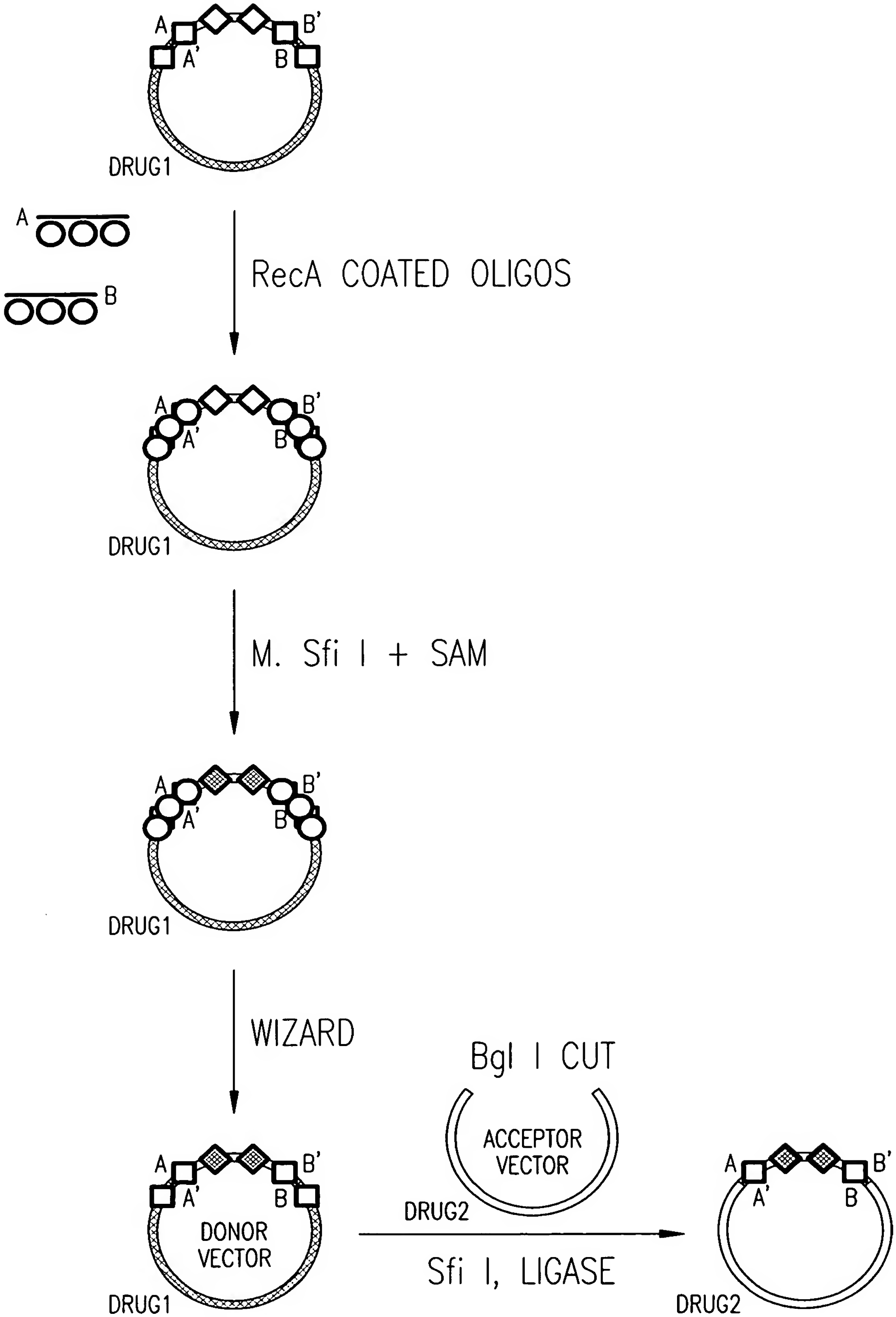
A
B'
A'
B
DRUG1
A
B
RecA COATED OLIGOS
A
B'
A'
B
DRUG1
M. Sfi I + SAM
A
B'
A'
B
DRUG1
WIZARD
Bgl I CUT
ACCEPTOR VECTOR
DRUG2
A
B'
A'
B
DONOR VECTOR
DRUG1
Sfi I, LIGASE
A
B'
A'
B
DRUG2

FIG. 10B

REs THAT CAN MAKE Sfi I ONE-WAY

3' 3b OVERHANG	RESTICTION ENZYME	RECOGNITION SEQUENCE
CNG	FmuI	G_GNC^C
CNG	PssI	RG_GNC^CY
CWG	Psp03I	G_GWC^C
GNC	BthCI	G_CNG^C
GSC	TauI	G_CSG^C
NNN	AlwNI	CAG_NNN^CTG
NNN	BglI	GCCN_NNN^NGGC
NNN	BsiYI	CCNN_NNN^NNGG
NNN	BstAPI	GCAN_NNN^NTGC
NNN	DraIII	CAC_NNN^GTG
NNN	MwoI	GCNN_NNN^NNGC
NNN	PflMI	CCAN_NNN^NTGG
NNN	RleAI	CCCACANNNNNNNNN_NNN^
NNN	SfiI	GGCCN_NNN^NGGCC

FIG. 11

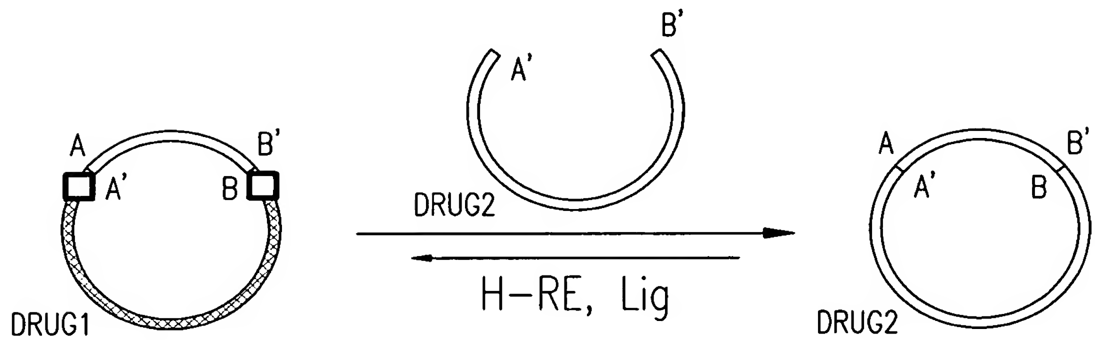

FIG. 12A

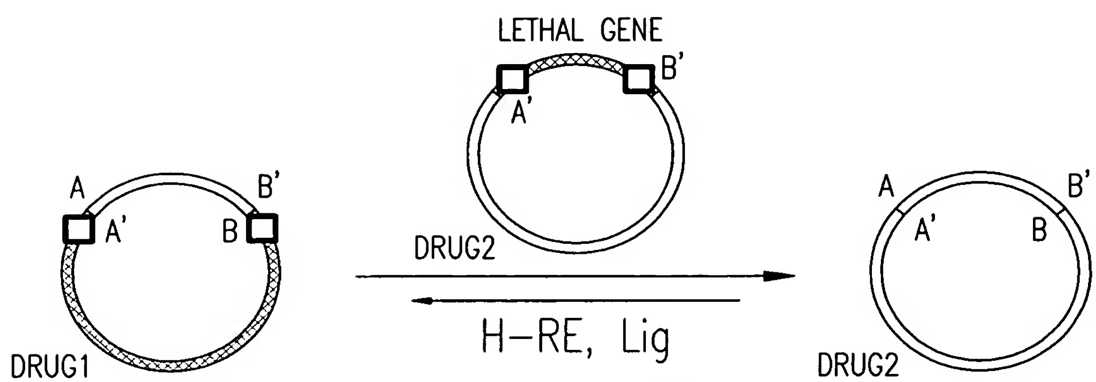

FIG. 12B

Sap I

- HOW TO MAKE Sap I "ONE WAY"
  - METHYLASES
  - ORIENTATION OF SITES IN VECTOR BACKBONE IN DONOR VECTOR AND IN ACCEPTOR VECTOR
  - LETHAL GENES IN STUFFER FRAGMENTS
  - Ear I, NOT Sap I SITES, IN ACCEPTOR VECTORS

G C T C T T C N^N N N
C G A G A A G N N N N^

C T C T T C N^N N N
G A G A A G N N N N^

- KEY ADVANTAGE OF Sap I
  - ONLY THREE BASES PER EXCHANGE SITE LEFT IN ACCEPTOR VECTOR

FIG. 13

TWO ENZYME APPROACH

- Sgf I -
  CUTTER OF HUMAN cDNAs, TWO BASE 3' OVERHANG

  G C G A T^C G C
  C G C^T A G C G

- Pme I -
  CUTTER, BLUNT END CUTTER

  G T T T^A A A C
  C A A A^T T T G

FIG. 14A

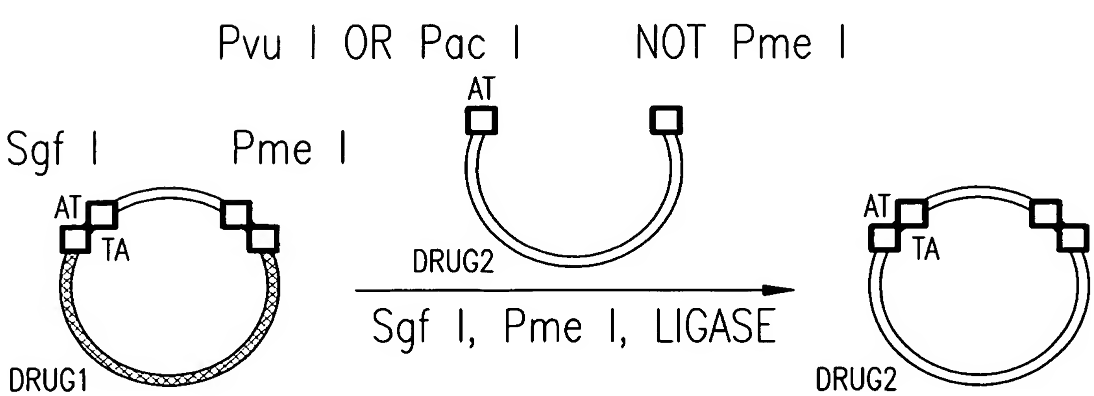

FIG. 14B

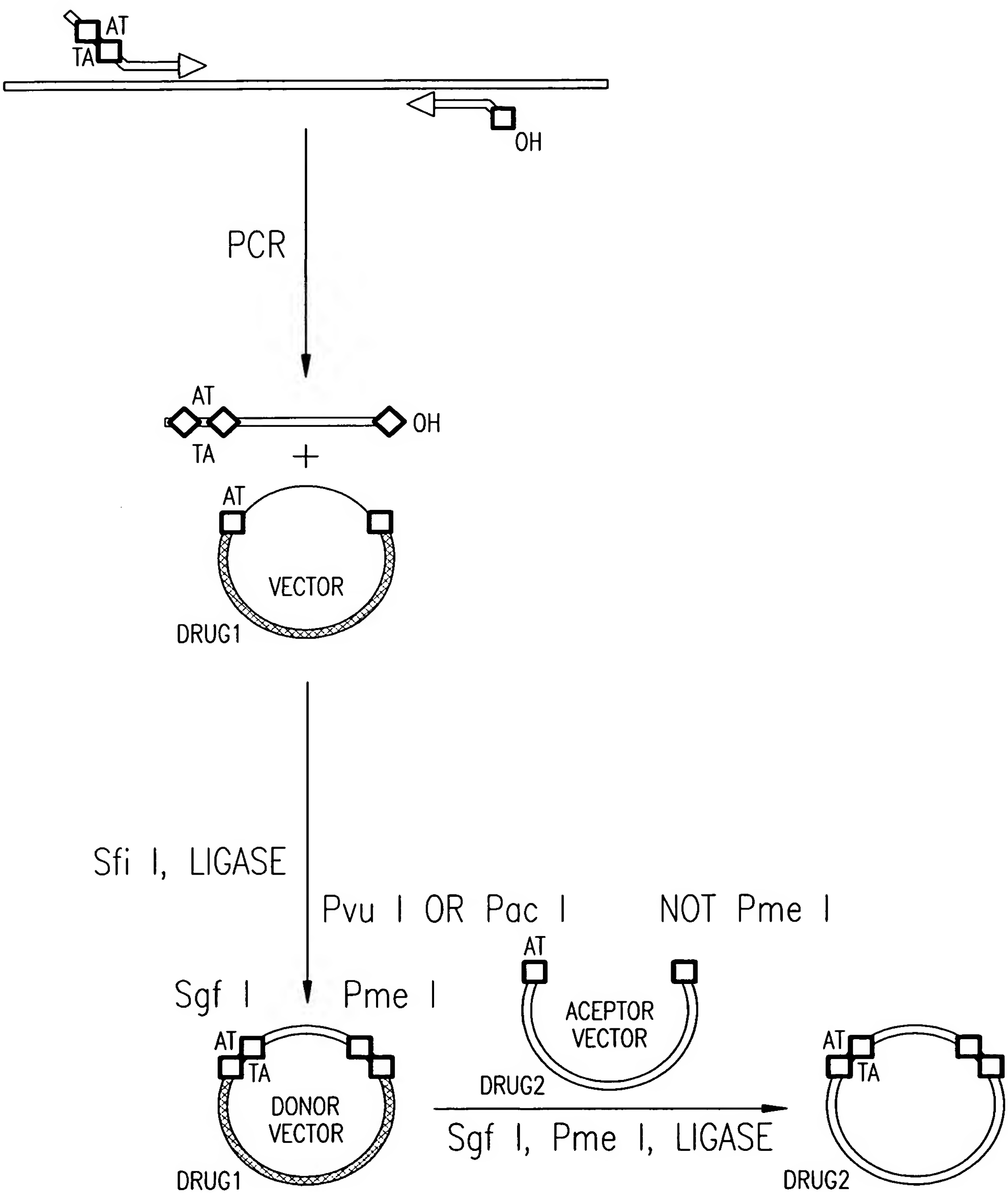
AT
TA
OH
PCR
AT
TA
OH
+
AT
VECTOR
DRUG1
Sfi I, LIGASE
Pvu I OR Pac I
NOT Pme I
Sgf I
Pme I
AT
ACEPTOR
VECTOR
DRUG2
AT
TA
DONOR
VECTOR
DRUG1
Sgf I, Pme I, LIGASE
AT
TA
DRUG2

FIG. 15

N-TERMINAL Sgf I SITE CAN ALLOW
N-TERMINAL FUSION OR NO FUSION

NAAGGAGCGATCGCCATGg
--RBS- Kozak--

VAAGGAGCGATCGCCATG
KEQGlyAlAIleAlaMet

FIG. 16

C-TERMINAL Pme I SITE ALLOWS TERMINATION
(+1AA) OR C-TERMINAL FUSIONS

NNNGTTTAAACN
XaaValTer

NNNGTTTATCN with EcoRV
XaaValTyr

NNNGTTTCCAN with BalI, etc.
XaaValSer

FIG. 17

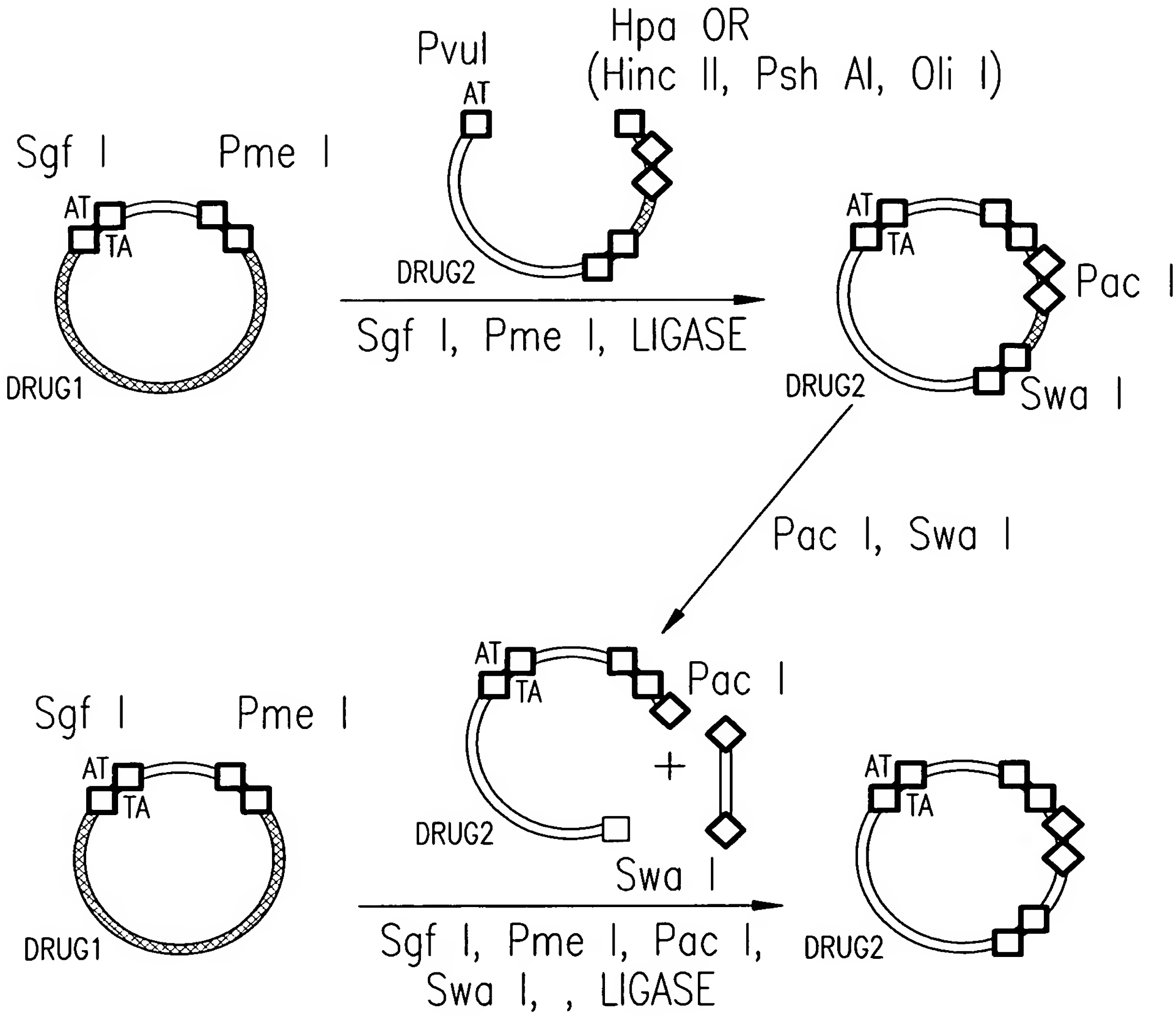
Sgf I
Pme I
AT
TA
DRUG1
PvuI
AT
Hpa OR
(Hinc II, Psh AI, Oli I)
DRUG2
Sgf I, Pme I, LIGASE
AT
TA
Pac I
DRUG2
Swa I
Pac I, Swa I
Sgf I
Pme I
AT
TA
DRUG1
AT
TA
Pac I
+
DRUG2
Swa I
Sgf I, Pme I, Pac I,
Swa I, , LIGASE
AT
TA
DRUG2

FIG. 18

N-TERMINAL Pac I--Sgf I FUSION SITE

NAAGGATTAATCGCCATGg
KEQGlyLeuIleAlaMet

C-TERMINAL Pme I--Swa I FUSION SITE

NNNGTTTAAATN
XaaValTer

FIG. 19A

N-TERMINAL Pac I--Sgf I FUSION SITE

NAAGGATTAATCGCCATGg
--RBS Kozak--

C-TERMINAL Pme I--Swa I FUSION SITE

NNNGTTTAAATN
XaaValTer

FIG. 19B

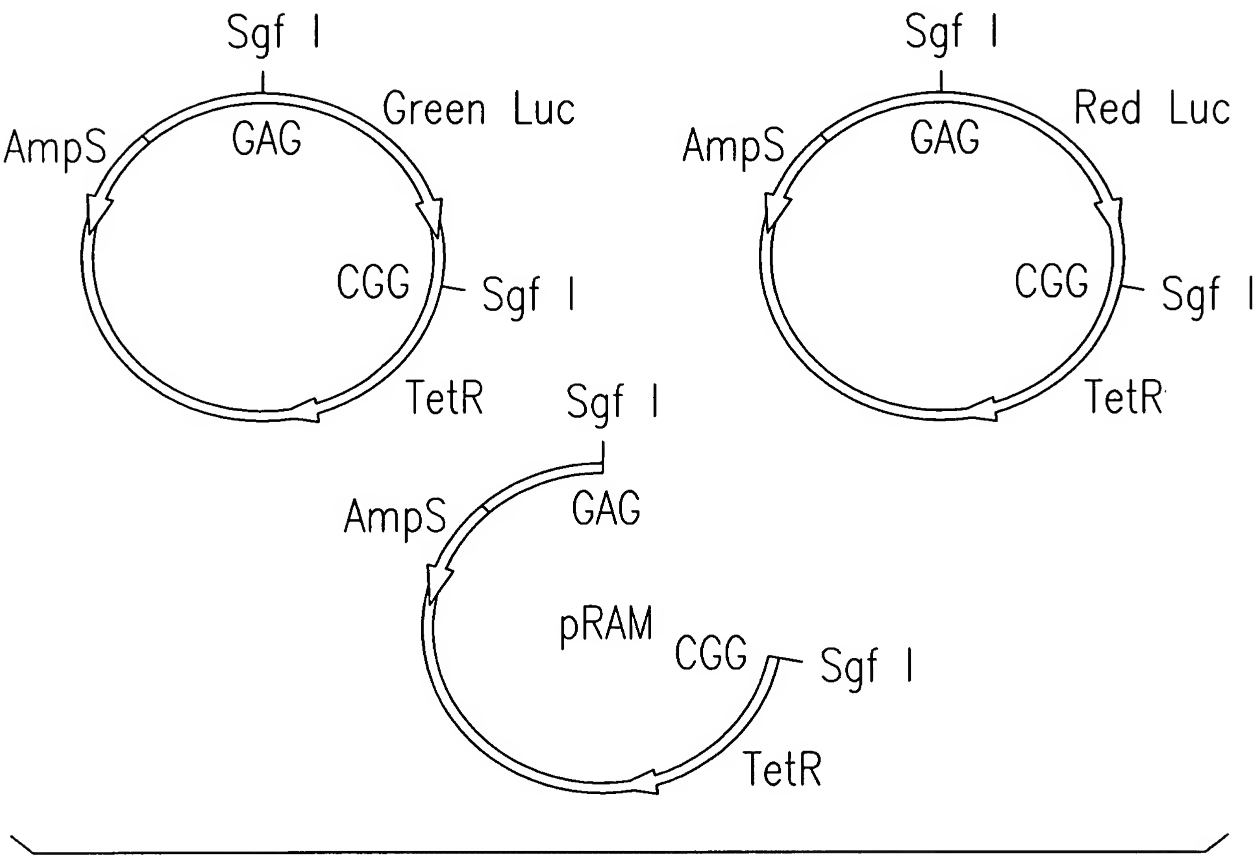
Sgf I
Green Luc
AmpS
GAG
CGG
Sgf I
TetR
Sgf I
Red Luc
AmpS
GAG
CGG
Sgf I
TetR
Sgf I
AmpS
GAG
pRAM
CGG
Sgf I
TetR

FIG. 20A

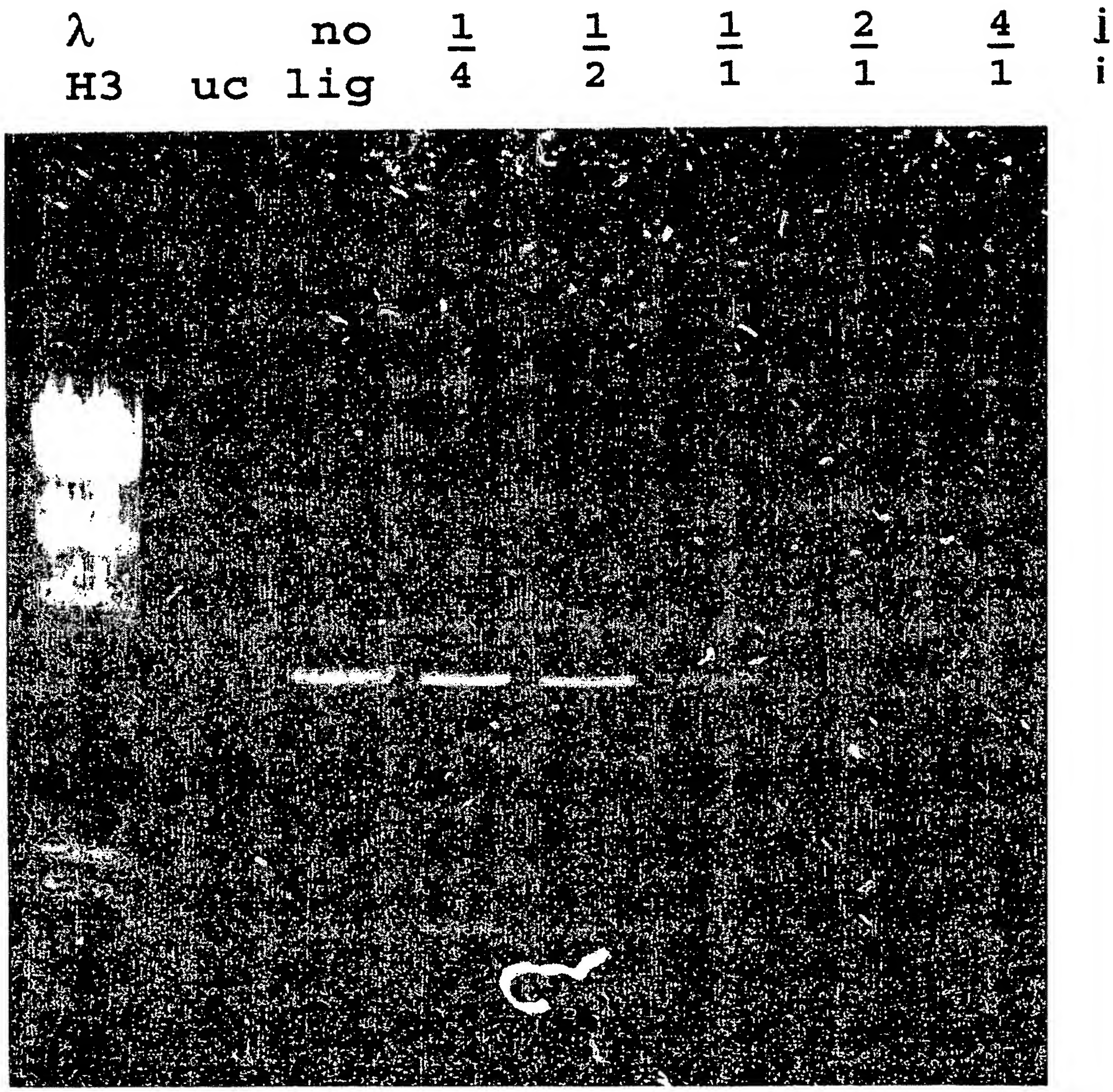

FIG. 20B

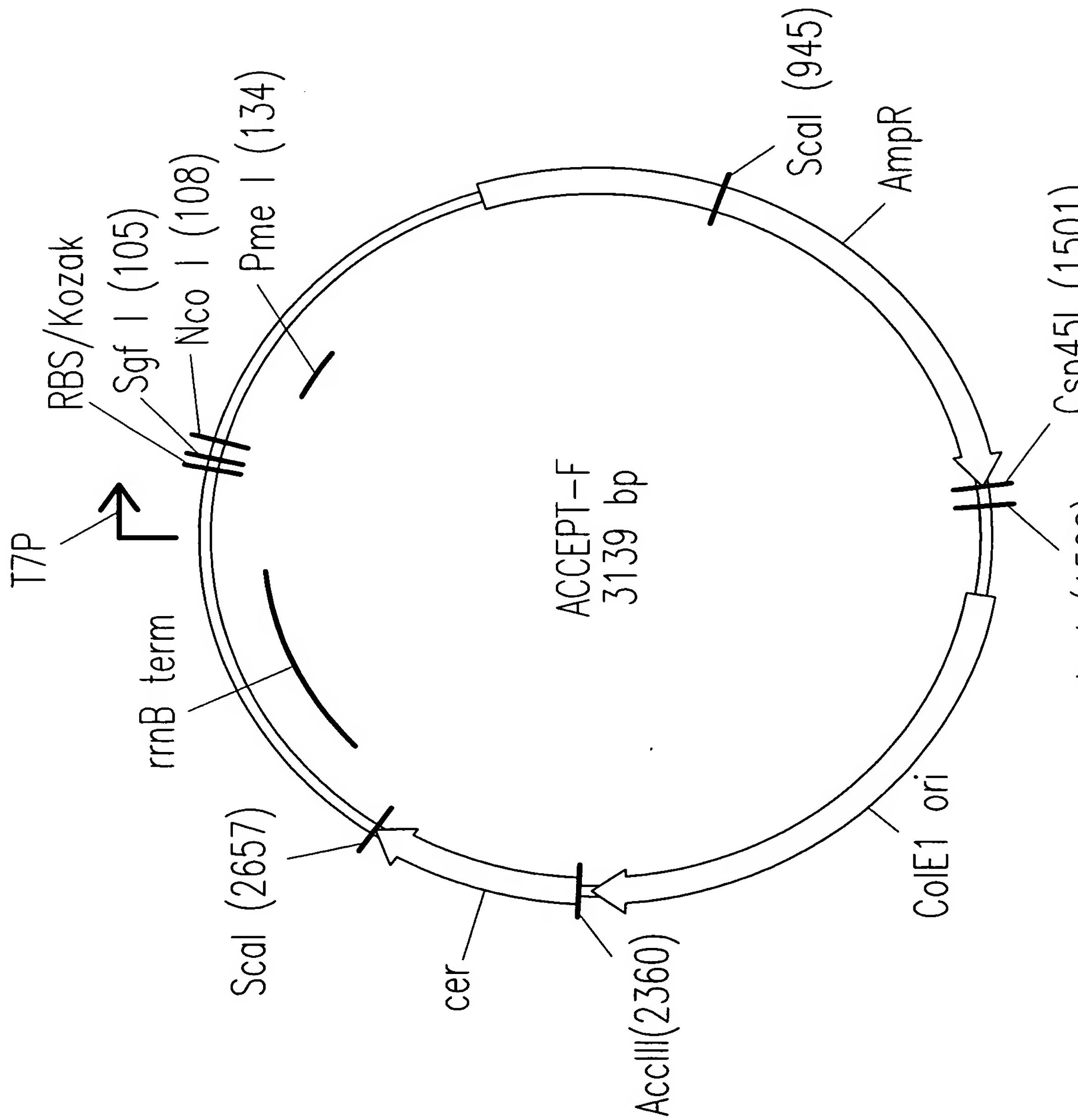
ACCEPT-F
3139 bp
T7P
RBS/Kozak
Sgf I (105)
Nco I (108)
Pme I (134)
ScaI (945)
AmpR
Csp45I (1501)
AgeI (1528)
ColE1 ori
AccIII(2360)
cer
ScaI (2657)
rrnB term

FIG. 21A

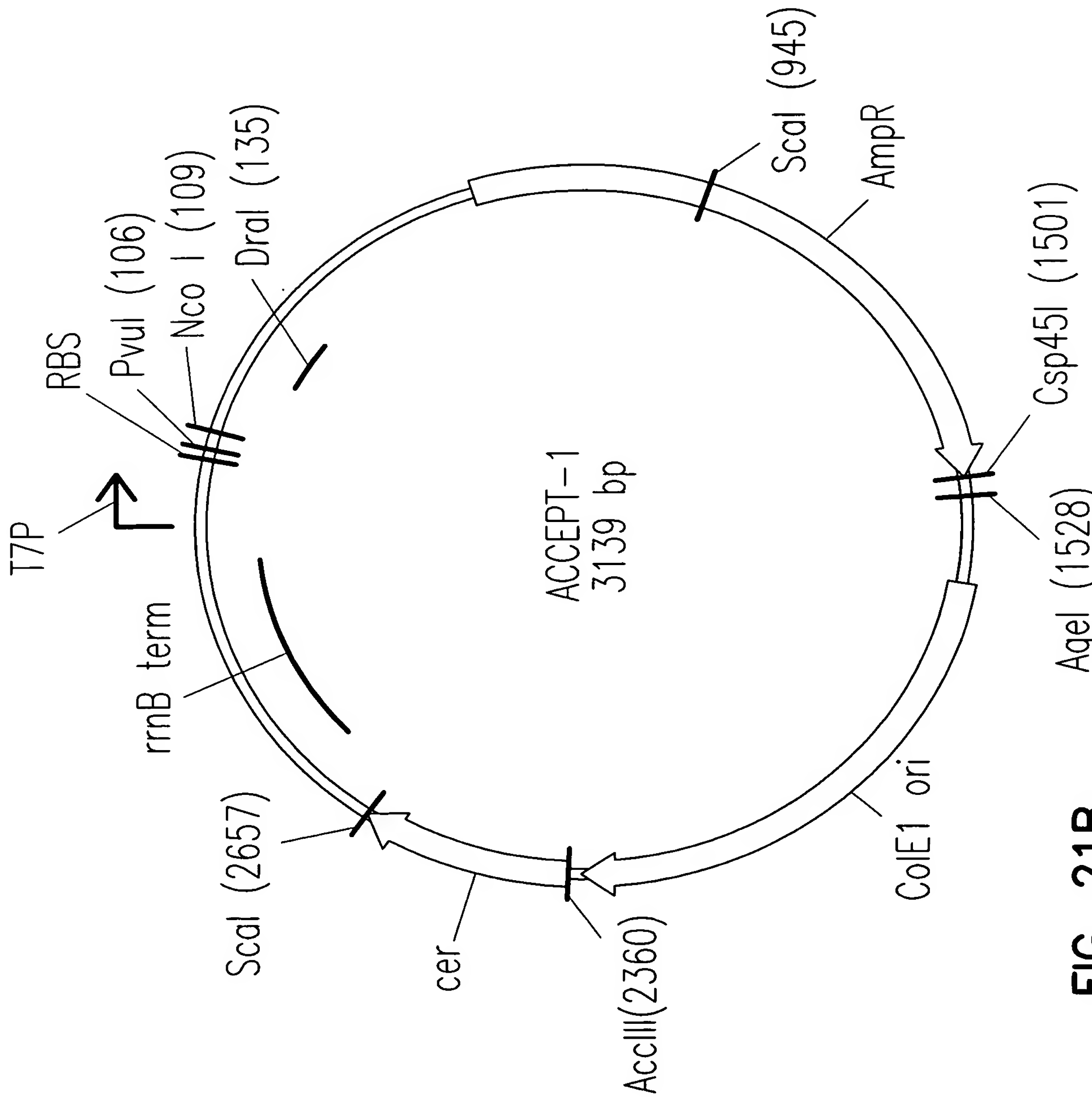
ACCEPT-1
3139 bp
T7P
RBS
PvuI (106)
Nco I (109)
DraI (135)
ScaI (945)
AmpR
Csp45I (1501)
AgeI (1528)
ColE1 ori
AccIII(2360)
cer
ScaI (2657)
rrnB term

FIG. 21B

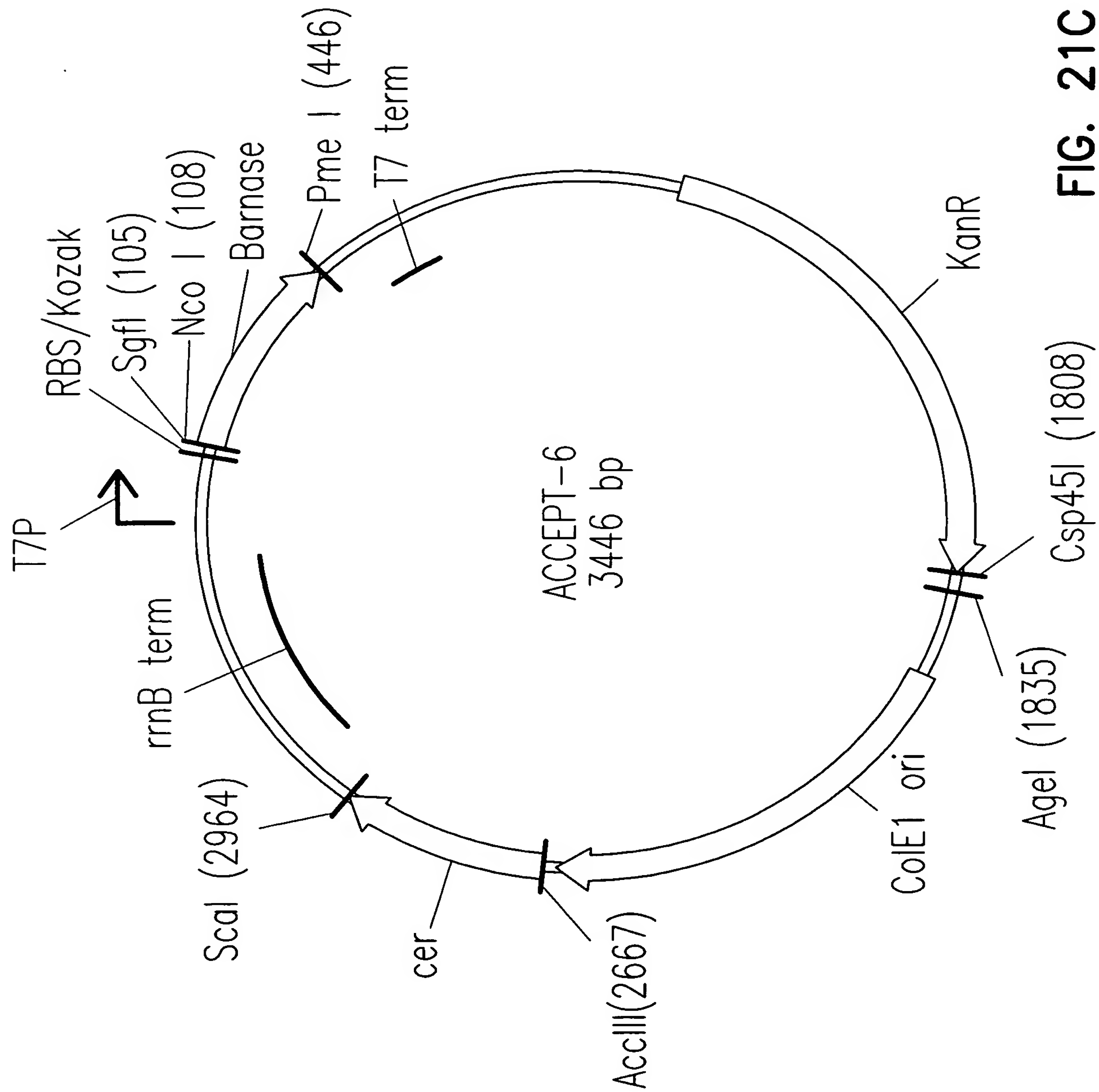
ACCEPT-6
3446 bp
T7P
RBS/Kozak
SgfI (105)
Nco I (108)
Barnase
Pme I (446)
T7 term
KanR
Csp45I (1808)
AgeI (1835)
ColE1 ori
AccIII(2667)
cer
ScaI (2964)
rrnB term

FIG. 21C

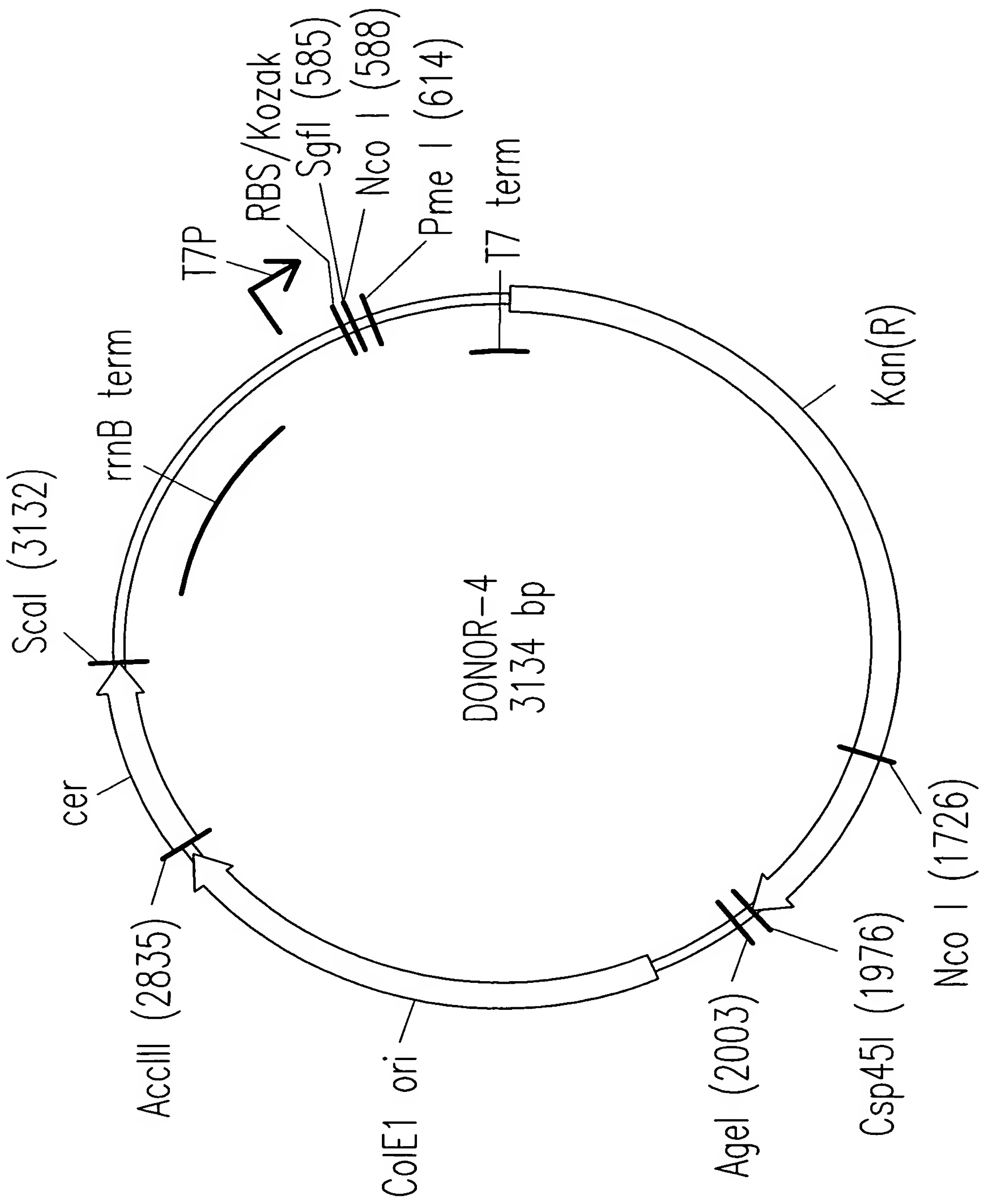
DONOR-4
3134 bp
T7P
RBS/Kozak
SgfI (585)
Nco I (588)
Pme I (614)
T7 term
Kan(R)
Nco I (1726)
Csp45I (1976)
AgeI (2003)
ColE1 ori
AccIII (2835)
cer
ScaI (3132)
rrnB term

FIG. 21D

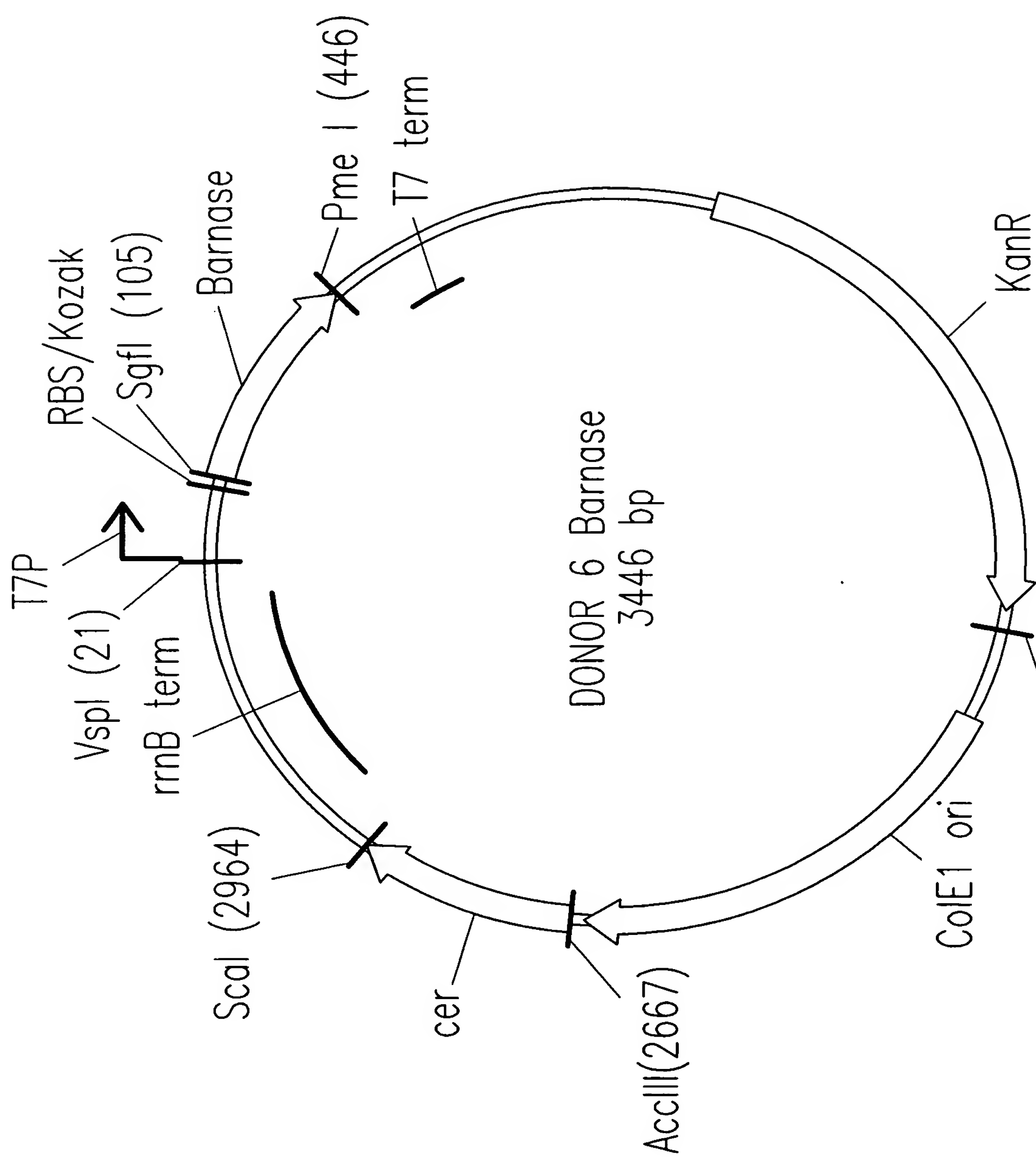
DONOR 6 Barnase
3446 bp
T7P
VspI (21)
rrnB term
RBS/Kozak
SgfI (105)
Barnase
Pme I (446)
T7 term
KanR
AgeI (1835)
ColE1 ori
AccIII(2667)
cer
ScaI (2964)

FIG. 21E

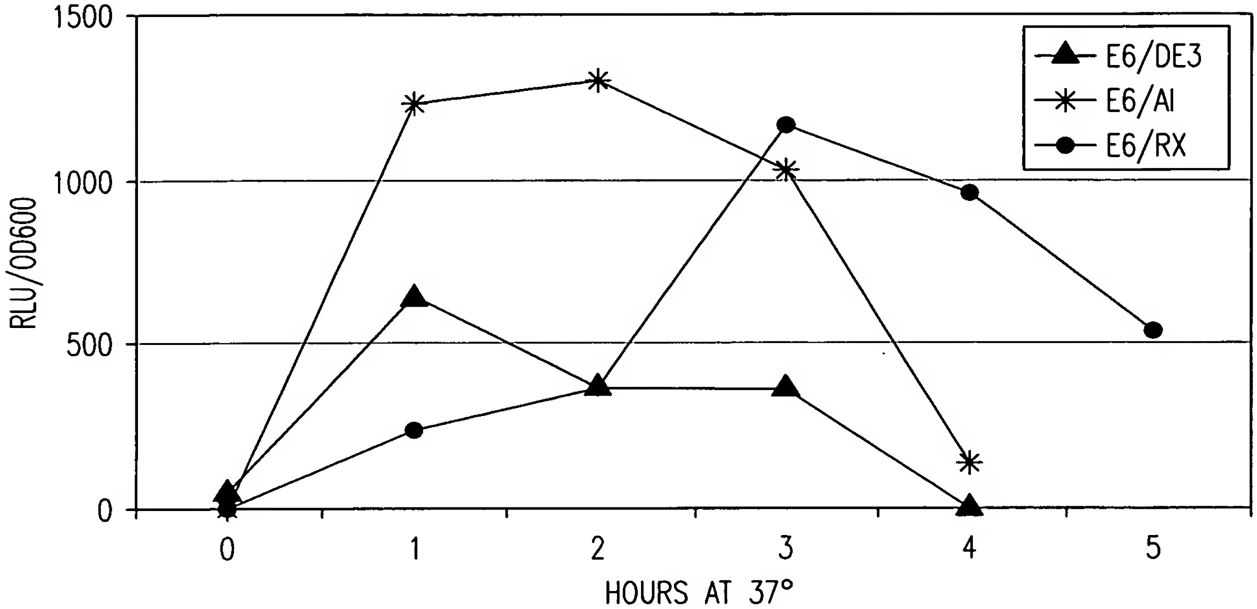

FIG. 22A

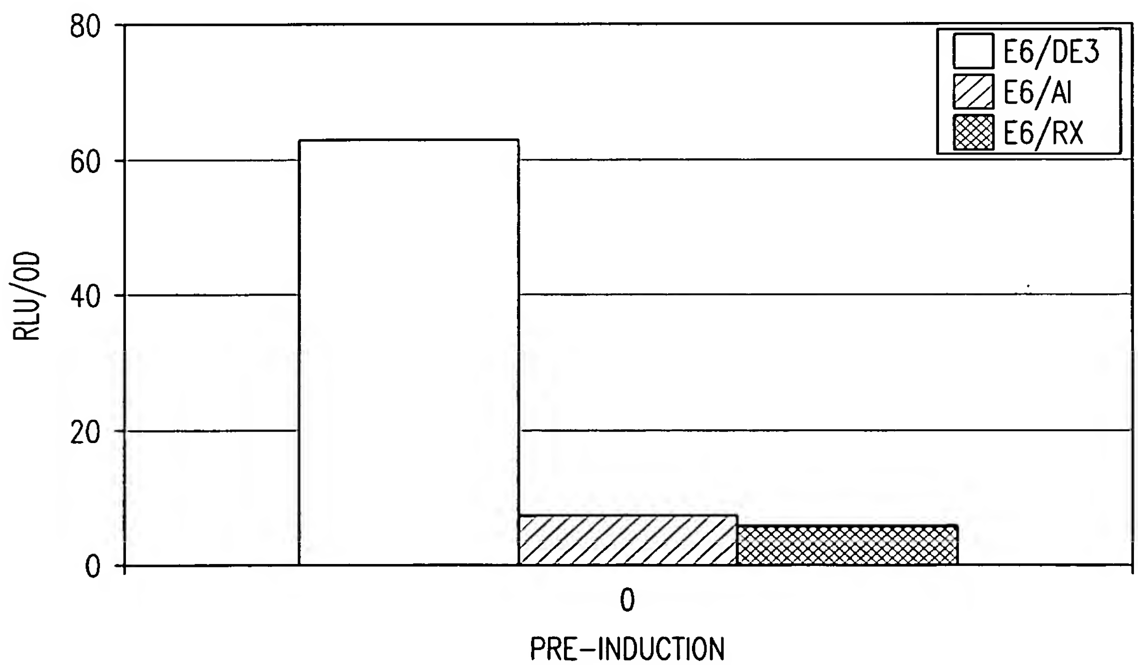

FIG. 22B